

American Mathematical Society

Colloquium Publications

Volume 60

Large Networks and Graph Limits

László Lovász



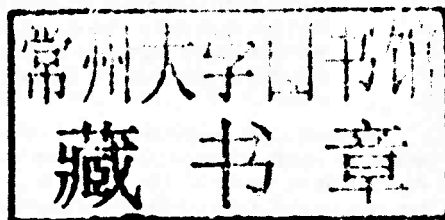
American Mathematical Society

Colloquium Publications

Volume 60

Large Networks and Graph Limits

László Lovász



American Mathematical Society
Providence, Rhode Island

Editorial Board

Lawrence C. Evans
Yuri Manin
Peter Sarnak (Chair)

2010 *Mathematics Subject Classification*. Primary 58J35, 58D17, 58B25, 19L64, 81R60, 19K56, 22E67, 32L25, 46L80, 17B69.

For additional information and updates on this book, visit
www.ams.org/bookpages/coll-60

ISBN-13: 978-0-8218-9085-1

Copying and reprinting. Individual readers of this publication, and nonprofit libraries acting for them, are permitted to make fair use of the material, such as to copy a chapter for use in teaching or research. Permission is granted to quote brief passages from this publication in reviews, provided the customary acknowledgment of the source is given.

Republication, systematic copying, or multiple reproduction of any material in this publication is permitted only under license from the American Mathematical Society. Requests for such permission should be addressed to the Acquisitions Department, American Mathematical Society, 201 Charles Street, Providence, Rhode Island 02904-2294 USA. Requests can also be made by e-mail to reprint-permission@ams.org.

© 2012 by the author. All rights reserved.
Printed in the United States of America.

∞ The paper used in this book is acid-free and falls within the guidelines established to ensure permanence and durability.
Visit the AMS home page at <http://www.ams.org/>

10 9 8 7 6 5 4 3 2 1 17 16 15 14 13 12

Large Networks and Graph Limits

To Kati
as all my books

Preface

Within a couple of months in 2003, in the Theory Group of Microsoft Research in Redmond, Washington, three questions were asked by three colleagues. Michael Freedman, who was working on some very interesting ideas to design a quantum computer based on methods of algebraic topology, wanted to know which graph parameters (functions on finite graphs) can be represented as partition functions of models from statistical physics. Jennifer Chayes, who was studying internet models, asked whether there was a notion of “limit distribution” for sequences of graphs (rather than for sequences of numbers). Vera T. Sós, a visitor from Budapest interested in the phenomenon of quasirandomness and its connections to the Regularity Lemma, suggested to generalize results about quasirandom graphs to multitype quasirandom graphs. It turned out that these questions were very closely related, and the ideas which we developed for the answers have motivated much of my research for the next years.

Jennifer’s question recalled some old results of mine characterizing graphs through homomorphism numbers, and another paper with Paul Erdős and Joel Spencer in which we studied normalized versions of homomorphism numbers and their limits. Using homomorphism numbers, Mike Freedman, Lex Schrijver and I found the answer to Mike’s question in a few months. The method of solution, the use of graph algebras, provided a tool to answer Vera’s. With Christian Borgs, Jennifer Chayes, Lex Schrijver, Vera Sós, Balázs Szegedy, and Kati Vesztegombi, we started to work out an algebraic theory of graph homomorphisms and an analytic theory of convergence of graph sequences and their limits. This book will try to give an account of where we stand.

Finding unexpected connections between the three questions above was stimulating and interesting, but soon we discovered that these methods and results are connected to many other studies in many branches of mathematics. A couple of years earlier Itai Benjamini and Oded Schramm had defined convergence of graph sequences with bounded degree, and constructed limit objects for them (our main interest was, at least initially, the convergence theory of dense graphs). Similar ideas were raised even earlier by David Aldous. The limit theories of dense and bounded-degree graphs have lead to many analogous questions and results, and each of them is better understood thanks to the other.

Statistical physics deals with very large graphs and their local and global properties, and it turned out to be extremely fruitful to have two statistical physicists (Jennifer and Christian) on the (informal) team along with graph theorists. This put the burden to understand the other person’s goals and approaches on all of us, but at the end it was the key to many of the results.

Another important connection that was soon discovered was the theory of property testing in computer science, initiated by Goldreich, Goldwasser and Ron several years earlier. This can be viewed as statistics done on graphs rather than on numbers, and probability and statistics became a major tool for us.

One of the most important application areas of these results is extremal graph theory. A fundamental tool in the extremal theory of dense graphs is Szemerédi's Regularity Lemma, and this lemma turned out to be crucial for us as well. Graph limit theory, we hope, repaid some of this debt, by providing the shortest and most general formulation of the Regularity Lemma ("compactness of the graphon space"). Perhaps the most exciting consequence of the new theory is that it allows the precise formulation of, and often the exact answer to, some very general questions concerning algorithms on large graphs and extremal graph theory. Independently and about the same time as we did, Razborov developed the closely related theory of flag algebras, which has led to the solution of several long-standing open problems in extremal graph theory.

Speaking about limits means, of course, analysis, and for some of us graph theorists, it meant hard work learning the necessary analytical tools (mostly measure theory and functional analysis, but even a bit of differential equations). Involving analysis has advantages even for some of the results that can be stated and proved purely graph-theoretically: many definitions and proofs are shorter, more transparent in the analytic language. Of course, combinatorial difficulties don't just disappear: sometimes they are replaced by analytic difficulties. Several of these are of a technical nature: Are the sets we consider Lebesgue/Borel measurable? In a definition involving an infimum, is it attained? Often this is not really relevant for the development of the theory. Quite often, on the other hand, measurability carries combinatorial meaning, which makes this relationship truly exciting.

There were some interesting connections with algebra too. Balázs Szegedy solved a problem that arose as a dual to the characterization of homomorphism functions, and through his proof he established, among others, a deep connection with the representation theory of algebras. This connection was later further developed by Schrijver and others. Another one of these generalizations has led to a combinatorial theory of categories, which, apart from some sporadic results, has not been studied before. The limit theory of bounded degree graphs also found very strong connections to algebra: finitely generated infinite groups yield, through their Cayley graphs, infinite bounded degree graphs, and representing these as limits of finite graphs has been studied in group theory (under the name of sofic groups) earlier.

These connections with very different parts of mathematics made it quite difficult to write this book in a readable form. One way out could have been to focus on graph theory, not to talk about issues whose motivation comes from outside graph theory, and sketch or omit proofs that rely on substantial mathematical tools from other parts. I felt that such an approach would hide what I found the most exciting feature of this theory, namely its rich connections with other parts of mathematics (classical and non-classical). So I decided to explain as many of these connections as I could fit in the book; the reader will probably skip several parts if he/she does not like them or does not have the appropriate background, but perhaps the flavor of these parts can be remembered.

The book has five main parts. First, an informal introduction to the mathematical challenges provided by large networks. We ask the “general questions” mentioned above, and try to give an informal answer, using relatively elementary mathematics, and motivating the need for those more advanced methods that are developed in the rest of the book.

The second part contains an algebraic treatment of homomorphism functions and other graph parameters. The two main algebraic constructions (connection matrices and graph algebras) will play an important role later as well, but they also shed some light on the seemingly completely heterogeneous set of “graph parameters”.

In the third part, which is the longest and perhaps most complete within its own scope, the theory of convergent sequences of dense graphs is developed, and applications to extremal graph theory and graph algorithms are given.

The fourth part contains an analogous theory of convergent sequences of graphs with bounded degree. This theory is more difficult and less well developed than the dense case, but it has even more important applications, not only because most networks arising in real life applications have low density, but also because of connections with the theory of finitely generated groups. Research on this topic has been perhaps the most active during the last months of my work, so the topic was a “moving target”, and it was here where I had the hardest time drawing the line where to stop with understanding and explaining new results.

The fifth part deals with extensions. One could try to develop a limit theory for almost any kind of finite structures. Making a somewhat arbitrary selection, we only discuss extensions to edge-coloring models and categories, and say a few words about hypergraphs, to much less depth than graphs are discussed in parts III and IV.

I included an Appendix about several diverse topics that are standard mathematics, but due to the broad nature of the connections of this material in mathematics, few readers would be familiar with all of them.

One of the factors that contributed to the (perhaps too large) size of this book was that I tried to work out many examples of graph parameters, graph sequences, limit objects, etc. Some of these may be trivial for some of the readers, others may be tough, depending on one’s background. Since this is the first monograph on the subject, I felt that such examples would help the reader to digest this quite diverse material.

In addition, I included quite a few exercises. It is a good trick to squeeze a lot of material into a book through this, but (honestly) I did try to find exercises about which I expected that, say, a graduate student of mathematics could solve them with not too much effort.

Acknowledgements. I am very grateful to my coauthors of those papers that form the basis of this book: Christian Borgs, Jennifer Chayes, Michael Freedman, Lex Schrijver, Vera Sós, Balázs Szegedy, and Kati Vesztegombi, for sharing their ideas, knowledge, and enthusiasm during our joint work, and for their advice and extremely useful criticism in connection with this book. The creative atmosphere and collaborative spirit at Microsoft Research made the successful start of this research project possible. It was a pleasure to do the last finishing touches on the book in Redmond again. The author acknowledges the support of ERC Grant No. 227701,

OTKA grant No. CNK 77780, the hospitality of the Institute for Advanced Study in Princeton, and in particular of Avi Wigderson, while writing most of this book.

My wife Kati Vesztergombi has not only contributed to the content, but has provided invaluable professional, technical and personal help all the time.

Many other colleagues have very unselfishly offered their expertise and advice during various phases of our research and while writing this book. I am particularly grateful to Miklós Abért, Noga Alon, Endre Csóka, Gábor Elek, Guus Regts, Svante Janson, Dávid Kunszenti-Kovács, Gábor Lippner, Russell Lyons, Jarik Nešetřil, Yuval Peres, Oleg Pikhurko, the late Oded Schramm, Miki Simonovits, Vera Sós, Kevin Walker, and Dominic Welsh. Without their interest, encouragement and help, I would not have been able to finish my work.

Contents

Preface	xi
Part 1. Large graphs: an informal introduction	1
Chapter 1. Very large networks	3
1.1. Huge networks everywhere	3
1.2. What to ask about them?	4
1.3. How to obtain information about them?	5
1.4. How to model them?	8
1.5. How to approximate them?	11
1.6. How to run algorithms on them?	18
1.7. Bounded degree graphs	22
Chapter 2. Large graphs in mathematics and physics	25
2.1. Extremal graph theory	25
2.2. Statistical physics	32
Part 2. The algebra of graph homomorphisms	35
Chapter 3. Notation and terminology	37
3.1. Basic notation	37
3.2. Graph theory	38
3.3. Operations on graphs	39
Chapter 4. Graph parameters and connection matrices	41
4.1. Graph parameters and graph properties	41
4.2. Connection matrices	42
4.3. Finite connection rank	45
Chapter 5. Graph homomorphisms	55
5.1. Existence of homomorphisms	55
5.2. Homomorphism numbers	56
5.3. What hom functions can express	62
5.4. Homomorphism and isomorphism	68
5.5. Independence of homomorphism functions	72
5.6. Characterizing homomorphism numbers	75
5.7. The structure of the homomorphism set	79
Chapter 6. Graph algebras and homomorphism functions	83
6.1. Algebras of quantum graphs	83
6.2. Reflection positivity	88

6.3. Contractors and connectors	94
6.4. Algebras for homomorphism functions	101
6.5. Computing parameters with finite connection rank	106
6.6. The polynomial method	108
Part 3. Limits of dense graph sequences	113
Chapter 7. Kernels and graphons	115
7.1. Kernels, graphons and stepfunctions	115
7.2. Generalizing homomorphisms	116
7.3. Weak isomorphism I	121
7.4. Sums and products	122
7.5. Kernel operators	124
Chapter 8. The cut distance	127
8.1. The cut distance of graphs	127
8.2. Cut norm and cut distance of kernels	131
8.3. Weak and L_1 -topologies	138
Chapter 9. Szemerédi partitions	141
9.1. Regularity Lemma for graphs	141
9.2. Regularity Lemma for kernels	144
9.3. Compactness of the graphon space	149
9.4. Fractional and integral overlays	151
9.5. Uniqueness of regularity partitions	154
Chapter 10. Sampling	157
10.1. W -random graphs	157
10.2. Sample concentration	158
10.3. Estimating the distance by sampling	160
10.4. The distance of a sample from the original	164
10.5. Counting Lemma	167
10.6. Inverse Counting Lemma	169
10.7. Weak isomorphism II	170
Chapter 11. Convergence of dense graph sequences	173
11.1. Sampling, homomorphism densities and cut distance	173
11.2. Random graphs as limit objects	174
11.3. The limit graphon	180
11.4. Proving convergence	185
11.5. Many disguises of graph limits	193
11.6. Convergence of spectra	194
11.7. Convergence in norm	196
11.8. First applications	197
Chapter 12. Convergence from the right	201
12.1. Homomorphisms to the right and multicuts	201
12.2. The overlay functional	205
12.3. Right-convergent graphon sequences	207
12.4. Right-convergent graph sequences	211

Chapter 13. On the structure of graphons	217
13.1. The general form of a graphon	217
13.2. Weak isomorphism III	220
13.3. Pure kernels	222
13.4. The topology of a graphon	225
13.5. Symmetries of graphons	234
Chapter 14. The space of graphons	239
14.1. Norms defined by graphs	239
14.2. Other norms on the kernel space	242
14.3. Closures of graph properties	247
14.4. Graphon varieties	250
14.5. Random graphons	256
14.6. Exponential random graph models	259
Chapter 15. Algorithms for large graphs and graphons	263
15.1. Parameter estimation	263
15.2. Distinguishing graph properties	266
15.3. Property testing	268
15.4. Computable structures	276
Chapter 16. Extremal theory of dense graphs	281
16.1. Nonnegativity of quantum graphs and reflection positivity	281
16.2. Variational calculus of graphons	283
16.3. Densities of complete graphs	285
16.4. The classical theory of extremal graphs	293
16.5. Local vs. global optima	294
16.6. Deciding inequalities between subgraph densities	299
16.7. Which graphs are extremal?	307
Chapter 17. Multigraphs and decorated graphs	317
17.1. Compact decorated graphs	318
17.2. Multigraphs with unbounded edge multiplicities	325
Part 4. Limits of bounded degree graphs	327
Chapter 18. Graphings	329
18.1. Borel graphs	329
18.2. Measure preserving graphs	332
18.3. Random rooted graphs	338
18.4. Subgraph densities in graphings	344
18.5. Local equivalence	346
18.6. Graphings and groups	349
Chapter 19. Convergence of bounded degree graphs	351
19.1. Local convergence and limit	351
19.2. Local-global convergence	360
Chapter 20. Right convergence of bounded degree graphs	367
20.1. Random homomorphisms to the right	367
20.2. Convergence from the right	375

Chapter 21. On the structure of graphings	383
21.1. Hyperfiniteness	383
21.2. Homogeneous decomposition	393
Chapter 22. Algorithms for bounded degree graphs	397
22.1. Estimable parameters	397
22.2. Testable properties	402
22.3. Computable structures	405
Part 5. Extensions: a brief survey	413
Chapter 23. Other combinatorial structures	415
23.1. Sparse (but not very sparse) graphs	415
23.2. Edge-coloring models	416
23.3. Hypergraphs	421
23.4. Categories	425
23.5. And more...	429
Appendix A. Appendix	433
A.1. Möbius functions	433
A.2. The Tutte polynomial	434
A.3. Some background in probability and measure theory	436
A.4. Moments and the moment problem	441
A.5. Ultraproduct and ultralimit	444
A.6. Vapnik–Chervonenkis dimension	445
A.7. Nonnegative polynomials	446
A.8. Categories	447
Bibliography	451
Author Index	465
Subject Index	469
Notation Index	473

Part 1

Large graphs: an informal introduction

CHAPTER 1

Very large networks

1.1. Huge networks everywhere

In the last decade it became apparent that a large number of the most interesting structures and phenomena of the world can be described by networks: often the system consists of discrete, well separable elements, with connections (or interactions) between certain pairs of them. To understand the behavior of the whole system, one has to study the behavior of the individual elements as well as the structure of the underlying network. Let us see some examples.

- Among very large networks, probably the best known and the most studied is the *internet*. Moreover, the internet (as the physical underlying network) gives rise to many other networks: the network of hyperlinks (web, logical internet), internet based social networks, distributed data bases, etc. The size of the internet is growing fast: currently the number of web pages may be 30 billion ($3 \cdot 10^{10}$) or more, and the number of interconnected devices is probably more than a billion. The graph theoretic structure of the internet determines, to a large degree, how communication protocols should be designed, how likely certain parts get jammed, how fast computer viruses spread etc.
- Social networks are basic objects of many studies in the area of sociology, history, epidemiology and economics. They are not necessarily formally established, like Facebook and other internet networks: The largest social network is the acquaintance graph of all living people, with about 7 billion nodes. The structure of this acquaintance graph determines, among others, how fast news, inventions, religions, diseases spread over the world, now and during history.
- Biology contributes ecological networks, networks of interactions between proteins, and the human brain, just to mention a few. The human brain, a network of neurons, is really large for its mass, having about a hundred billion nodes. One of the greatest challenges is, of course, to understand ourselves.
- Statistical physics studies the interactions between large numbers of discrete particles, where the underlying structure is often described by a graph. For example, a crystal can be thought of as a graph whose nodes are the atoms and whose edges represent chemical bonds. A perfect crystal is a rather boring graph, but impurities and imperfections create interesting graph-theoretical digressions. 10 gram of a diamond has about 5×10^{23} nodes. The structure of a crystal influences important macroscopic properties like whether the material is magnetizable, or how it melts.
- Some of the largest networks in engineering occur in chip design. There can be more than a billion transistors on a chip nowadays. Even though these networks are man-made and carefully designed, many of their properties, like

the exact time they will need to perform some computation, are difficult to determine from their design, due to their huge size.

- To be pretentious, we can say that the whole universe is a single (really huge, possibly infinite) network, where the nodes are events (interactions between elementary particles), and the edges are the particles themselves. This is a network with perhaps 10^{80} nodes. It is an ongoing debate in physics how much additional structure the universe has, but perhaps understanding the graph-theoretical structure of this graph can help with understanding the global structure of the universe.

These huge networks pose exciting challenges for the mathematician. Graph Theory (the mathematical theory of networks) has been one of the fastest developing areas of mathematics in the last decades; with the appearance of the Internet, however, it faces fairly novel, unconventional problems. In traditional graph theoretical problems the whole graph is exactly given, and we are looking for relationships between its parameters or efficient algorithms for computing its parameters. On the other hand, very large networks (like the Internet) are never completely known, in most cases they are not even well defined. Data about them can be collected only by indirect means like random local sampling or by monitoring the behavior of various global processes.

Dense networks (in which a node is adjacent to a positive percent of other nodes) and very sparse networks (in which a node has a bounded number of neighbors) show a very different behavior. From a practical point of view, sparse networks are more important, but at present we have more complete theoretical results for dense networks. In this introduction, most of the discussion will focus on dense graphs; we will survey the additional challenges posed by sparse networks in Section 1.7.

1.2. What to ask about them?

Think of a really large graph, say the internet, and try to answer the following four simple questions about it.

Question 1. *Does the graph have an odd or even number of nodes?*

This is a very basic property of a graph in the classical setting. For example, it is one of the first theorems or exercises in a graph theory course that every graph with an odd number of nodes must have a node with even degree.

But for the internet, this question is clearly nonsense. Not only does the number of nodes change all the time, with devices going online and offline, but even if we fix a specific time like 12:00am today, it is not well-defined: there will be computers just in the process of booting up, breaking down etc.

Question 2. *What is the average degree of nodes?*

This, on the other hand, is a meaningful question. Of course, the average degree can only be determined with a certain error, and it will change as technology or the social composition of users change; but at a given time, a good approximation can be sought (I am not speaking now about how to find it).

Question 3. *Is the graph connected?*

To this question, the answer is almost certainly no: somewhere in the world there will be a faulty router with some unhappy users on the wrong side of it. But this is not the interesting way to interpret the question: we should consider the